Bacterial defenses and their trade-off with growth are not ubiquitous but depend on ecological contexts

Short title: Bacterial growth and defense

Zhi-Ling Liu[1,2], Jia Liu[1], Deng-Ke Niu[1,2*]

[1]MOE Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, 100875, China

[2]Beijing Key Laboratory of Gene Resource and Molecular Development, College of Life Sciences, Beijing Normal University, Beijing, 100875, China

*Correspondence author:

Deng-Ke Niu, College of Life Sciences, Beijing Normal University, Beijing, 100875, China.

dkniu@bnu.edu.cn, dengkeniu@hotmail.com

ORCID:

Zhi-Ling Liu, https://orcid.org/0000-0001-6579-6642

Deng-Ke Niu, https://orcid.org/0000-0003-4503-2658

Competing interests

The authors declare no competing interests.

Abstract

Bacteriophages, also known as bacterial viruses, exert significant influence on microbial ecosystems due to their widespread presence. In response to the threat of phage predation, bacteria have developed an extensive array of antiviral defense mechanisms. This study investigates the intricate relationship between bacterial defense mechanisms and growth rates across various ecological contexts. First, we found that bacteria lacking defense systems show prolonged doubling times. By integrating phylogenetic eigenvectors into the ecological feature matrix, we employed a phylogenetic random forest model to determine the most influential ecological features on the presence and abundance of bacterial defenses. Further phylogenetic regressions unveiled nuanced dependencies of bacterial defenses on specific environmental factors. Contrary to conventional assumptions, our findings challenge the idea of a universal distribution of bacterial defense systems, instead highlighting their reliance on subtle ecological characteristics. Particularly noteworthy are observations that symbiotic and endosymbiotic bacteria often exhibit reduced defense systems and negative correlations between minimal doubling time and defense system abundance. Conversely, as per conventional assumptions, free-living and motile bacteria displayed significant positive correlations between minimal doubling time and defense system abundance. Overall, our results suggest that the trade-off between growth and defense strategies is not universally applicable but rather contingent on context. Moreover, our study illuminates the critical role of ecological variables such as habitat specificity, motility, and nutrient availability in shaping bacterial growth rates and defense mechanisms. These findings underscore the complexity of microbial interactions and stress the importance of considering the ecological context in understanding bacterial defense strategies.

Keywords

Defense system, minimal doubling time, trade-off, trade-up, symbiosis, endosymbiosis, gut microbiome.

Introduction

Bacterial viruses, also known as phages or bacteriophages, represent the most abundant biological entities on Earth. They significantly outnumber bacterial cells across various samples and microbiomes, with ratios reaching up to ~15-25 fold, as observed in freshwater plankton (López-García et al. 2023; Roux and Brum 2023). With few exceptions, such as intracellular parasitic and symbiotic bacteria, the majority of bacterial species fall prey to phages (Koonin et al. 2017). In environments like seawater, this constant predation results in a substantial daily death rate of bacterioplankton, estimated between 25-40% (Wommack and Colwell 2000; Suttle 2007). To combat phage attacks, bacteria have evolved an arsenal of defense tactics targeting every stage of the phage's life cycle. These tactics include diverse mechanistic approaches such as the restriction-modification (RM) system, the Clustered Regularly Interspaced Short Palindromic Repeat and associated protein (CRISPR-Cas) system, and the Abortive Infection (Abi) system (Stern and Sorek 2011; Koonin et al. 2017; Bernheim and Sorek 2020), collectively referred to as antiviral defense systems.

Despite the potential for horizontal gene transfer, bacteria typically only possess a subset of the available anti-phage mechanisms (Rostøl and Marraffini 2019). For instance, the CRISPR-Cas systems, which provide adaptive immunity by targeting previously encountered pathogens, are found in approximately 40% of bacteria (Makarova et al. 2020). Similarly, Abi systems are sporadically distributed across bacterial phylogeny (Stern and Sorek 2011), while RM systems are widespread but absent in about 4% of bacterial genomes (Roberts et al. 2015). Additionally, the uneven distribution of antiviral defense systems extends to the archaeal lineage (Roberts et al. 2015; Makarova et al. 2020), reflecting trade-offs between the costs of maintaining such systems and the benefits gained in phage resistance (Winter et al. 2010; Ferenci 2016; Pančić and Kiørboe 2018; Rostøl and Marraffini 2019; Thingstad 2022; Liu et al. 2023).

In biological systems, traits often exhibit trade-offs due to constraints on resources such as energy, habitat, or time (Roff and Fairbairn 2007; Garland Jr 2014; Garland et al. 2022). For example, energy allocated for growth cannot simultaneously be used for defense or other energy-intensive processes, leading to necessary compromises among traits. While the concept of growth-defense

trade-offs is well-established in plants and animals, its relevance in virus-infected prokaryotes is less recognized (Herms and Mattson 1992; Ferenci 2016).

Laboratory research on phage resistance has extensively explored trade-offs using model bacteria and phages such as *Escherichia coli* and *Pseudomonas aeruginosa*. Phage resistance evolution significantly impacts bacterial fitness components, including slowed cell division rates, reduced nutrient uptake, decreased virulence, and impaired cell motility (Lenski 1988; Bohannan and Lenski 2000; Bohannan et al. 2002; Buckling et al. 2005; Lennon et al. 2007; Jessup and Bohannan 2008; Quance and Travisano 2009; Hall et al. 2012; Ferenci 2016; Burmeister and Turner 2020; Castledine et al. 2022). Notably, some phage-resistant strains exhibit trade-ups, where selection improves multiple fitness components simultaneously, including antibiotic resistance (Moulton-Brown and Friman 2018; Burmeister et al. 2020).

However, resistance to phages does not consistently correlate negatively with bacterial growth rates in natural environments. Studies examining marine *Synechococcus* strains and phage isolates, as well as bacteria and phages from horse chestnut leaves, have shown conflicting results regarding the growth-defense trade-off (Lennon et al. 2007; Koskella et al. 2011). Recent research on marine bacterial communities has observed a negative correlation between growth rates and resistance to protists and viruses, supporting the growth-defense trade-off hypothesis (Yang et al. 2023).

To gain a comprehensive understanding of the trade-off between prokaryotic growth and defense, investigations encompassing diverse environments and species are essential. Recent research analyzing growth rates and adaptive immunity across bacterial domains revealed a significant positive correlation, indicating evolutionary trade-offs between growth and adaptive defense against virulent phages (Liu et al., 2023). Additionally, besides CRISPR-Cas systems, prokaryotes possess numerous other antiviral mechanisms, such as the RM and Abi systems, warranting further exploration (Koonin et al., 2017; Bernheim and Sorek, 2020; Georjon and Bernheim, 2023; Tesson and Bernheim, 2023). Understanding the relationship between growth and antiviral defense across different environments and prokaryotic lineages holds promise for advancing our knowledge of evolution and informing the development of antimicrobial strategies.

To comprehend the trade-off between prokaryotic growth and defense, it's imperative to adopt

a holistic perspective. Despite occasional inconsistencies, our aim is to attain a global understanding of whether phage-resistant lineages generally exhibit slower growth rates compared to phage-sensitive ones across bacterial domains. Such insights could profoundly influence our comprehension of evolutionary processes and facilitate the development of novel antimicrobial strategies.

Recently, we conducted an investigation into the relationship between growth rate and adaptive immunity in 4142 bacteria, utilizing predicted minimal doubling times retrieved from the EGGO database and CRISPR-Cas systems annotated with CRISPRCasFinder (Couvin et al. 2018; Weissman et al. 2021; Liu et al. 2023). Our findings unveiled a significant positive correlation, indicating evolutionary trade-offs between growth and adaptive defense against virulent phages within the bacterial domain.

In addition to CRISPR-Cas systems, prokaryotes have evolved various other antiviral mechanisms, including the RM system, the Abi system, and numerous others that warrant further exploration (Koonin et al. 2017; Bernheim and Sorek 2020; Georjon and Bernheim 2023; Tesson and Bernheim 2023). Notably, certain prokaryotes, such as the delta proteobacteria *Desulfonema limicola*, can harbor over 50 defense systems in their genomes (Tesson et al. 2022). Recently, Tesson et al. (2022) developed DefenseFinder, a novel antivirus-system detection program based on the HMM mapping algorithm, which enables the identification of known antiviral systems across 21,000 fully sequenced prokaryotic genomes. Leveraging this tool, we explored the relationship between growth and antiviral defense across bacterial domains, as well as within specific environments.

Materials and Methods

The minimal doubling times of bacteria and archaea were extracted from the EGGO database (accessed on 2021-11-26), encompassing over 200,000 predictions of estimated growth rates (Weissman et al. 2021). These values underwent a log transformation using the natural logarithm as the base. Due to its smaller sample size, we opted not to utilize the dataset of prokaryotic empirical minimal doubling times provided by Madin et al. (2020) in this study. While empirical data typically offer higher accuracy compared to computationally generated predictions, predictive methodologies

have their own advantages (Liu et al. 2023). Empirical minimal doubling periods may offer superior advantages over computationally generated predictions due to the inherent limitations of predictive methodologies achieving 100% precision. However, the wide spectrum of genetic diversity observed among strains of individual bacterial species necessitates empirical data encompassing various factors, all procured from equivalent strains, a requirement that is not consistently met. Predicted phenotype values, including defense systems and growth rate, for a strain originate from the genome sequences of that very strain, potentially circumventing faults stemming from polymorphism within each species.

The antiviral defense systems of over 21,000 complete microbial genomes were sourced from Tesson et al. (2022). Phylogenetic relationships among the analyzed prokaryotes were obtained from the Genome Taxonomy Database (GTDB) (Parks et al., 2022) (accessed 2022-04-08). GTDB offers superior taxonomic resolution and more precise classification compared to traditional phylogenetic trees by leveraging genomic data.

The intersecting samples among the three aforementioned datasets consisted of 3532 bacterial genomes and 237 archaeal genomes, some of which belonged to the same species. However, upon intersecting with subsequent datasets, only a minimal number of archaeal species were acquired, rendering most analyses impractical. Consequently, this paper focuses solely on the study of bacteria.

Ecological characteristics of prokaryotes were obtained from the ProTraits database (Brbić et al. 2016), accessed on October 2, 2022, via "ProTraits_precisionScores.txt" available at http://protraits.irb.hr/data.html. This comprehensive database integrates data from various repositories such as GOLD, NCBI, and KEGG, utilizing automated text mining and comparative genomic techniques to associate microorganisms with phenotypes. Our study focuses on ecological features potentially influencing defense system distribution, excluding physiological traits like glycerol and glutamate. Additionally, features with ambiguous biological significance, automated physiological features from machine learning, and those with missing *P* values exceeding 10% of the overall sample were discarded, retaining 125 prominent ecological features for analysis. Each ecological feature in the database is represented by two columns, providing precision scores for both the presence and absence of the feature, respectively. To facilitate analysis, these scores were converted into an overall maximal confidence score ranging from 0 to 1, following the methodology of Weissman et al. (2019).

Missing values in the ProTraits database were imputed using phylogenetic information (Debastiani et al. 2021). The PVRdecomp function of the R package PVR (Version 0.3) (Santos 2018) was employed to extract eigenvectors from the phylogenetic distance matrix. Subsequently, the missForest function of the R package missForest (Version 1.5) (Stekhoven and Bühlmann 2011) was utilized to estimate missing values.

Random forest, an integrated algorithm of decision trees, was utilized for regression tasks in this study. This method optimizes regression tasks, effectively handles high dimensionality, mitigates overfitting risks, accommodates both categorical and numerical variables, and consistently delivers reliable outcomes, making it superior to alternative techniques. The eigenvectors extracted from the phylogenetic distance matrix were incorporated into the ecological feature matrix to control phylogenetic relationships within the random forest analysis. The randomForest package (version 4.7-1.1) (Liaw and Wiener 2002) was implemented in R (version 4.2.0).

Phylogenetic generalized least-squares (PGLS) analysis of correlations was conducted using the R package phylolm (version 2.6.2) (Ho and Ane 2014). The Pagel λ model was employed due to its generally lower AIC values in our dataset. Differences in minimal doubling times between the two groups were compared using the phylANOVA function (Garland et al. 1993), while phylogenetic logistic regression analysis with dichotomous variables as dependent variables was performed using the phyloglm function (Ives and Garland 2010). When multiple correlation analyses were performed on the same data, *P* values were corrected using the Benjamini & Hochberg method via the p.adjust function in R (Version 4.0.2).

Results

Phylogenetic Signals in Analyzed Features and Their Implications for Comparative Analysis

Due to shared ancestry among bacterial samples, feature values may exhibit non-independence, violating fundamental assumptions of conventional statistical methods. As such, using phylogenetic comparative methods becomes imperative to mitigate these effects (Felsenstein 1985; Symonds and

Blomberg 2014). We evaluated the phylogenetic signals of all analyzed features and observed significant signals in nearly all cases, with minimal doubling time demonstrating the strongest signal (Supplementary Table 1). Consequently, we will employ phylogenetic comparative methods whenever applicable.

A recent study by Chen et al. (2023) highlighted the potential for conflicting outcomes in PGLS analysis when interchanging dependent and independent variables. While PGLS cannot establish causality but rather identifies correlations, Chen et al. utilized simulated data and a gold standard to establish criteria for selecting the dependent variable. Their findings suggested that choosing the feature with a stronger phylogenetic signal as the dependent variable led to more accurate correlation detection than randomly selected variables. In line with this approach, we adopted the feature exhibiting a stronger phylogenetic signal as the dependent variable in all PGLS analyses conducted in this study.

Ecological Factors Influencing the Presence of Antiviral Defense Systems in Bacteria

Our analysis stems from a comprehensive program capable of detecting all known antiviral systems in prokaryotic genomes (Tesson et al. 2022). However, among the 3532 bacterial genomes examined, 7% (243) were found to lack any defense systems. Phylogenetic ANOVA analysis unveiled that these bacteria exhibited significantly longer minimal doubling times compared to those encoding one or more defense systems (Mean ± SD: 5.6 ± 3.1 h *vs*. 3.2 ± 4.4 h. $P = 0.0011$). Surprisingly, bacteria devoid of defense systems did not demonstrate faster proliferation rates, contradicting the anticipated trade-off between growth and defense. It is plausible that these bacteria inhabit specific ecological niches where neither defense mechanisms nor rapid growth are necessary.

To delve into the factors influencing the presence or absence of bacterial defense systems, we leveraged ecological data from the ProTraits database. Since the database primarily comprises bacterial species rather than individual strains or genomes, we aggregated some of the 3532 genomes into species. To capture species-level characteristics, we computed median values of defense system abundance and minimal doubling time for genomes within the same species, identifying 1044 distinct

bacterial species.

Utilizing a random forest model, we ranked the importance of 125 ecological features. Remarkably, factors such as "known habitats = insect endosymbiont," "known habitats = endosymbiont," "biotic relationship = symbiont," and "host = insects general" emerged as top influencers, underscoring the significance of symbiotic and endosymbiotic relationships (Fig. 1).

Given that the random forest model cannot discern the directional effect of features on the presence or absence of antiviral defense systems, we conducted phylogenetic logistic regression (Table 1). Results indicated a negative correlation between antiviral defense presence and "known habitats = insect endosymbiont," "biotic relationship = symbiont," and "host = insects general" ($P < 10^{-8}$ for all), while "known habitats = endosymbiont" showed no significant correlation (P = 0.452).

Ecological Determinants of Defense System Abundance: Insights from Symbiosis and Environmental Dynamics

The preceding study explored factors influencing the presence or absence of defense systems. Here, we delve deeper into the variables shaping defense system abundance. Bacteria lacking defense mechanisms will be recorded as having zero systems, ensuring inclusivity in our analysis.

Among 125 ecological features, "known habitats = insect endosymbiont," "known habitats = endosymbiont," and "biotic relationship = symbiont" emerged as top influencers, underscoring the importance of symbiotic associations (Fig. 2). Furthermore, symbiotic relationships with hosts were supported by additional features within the top rankings, including "ecosystem type = circulatory system," "ecosystem subtype = blood," "habitat = host associated," "host = insects general," "ecosystem category = mammals," "known habitats = host," and "ecosystem = host associated." If symbiotic bacteria tend to have fewer defense systems for certain reasons, free-living bacteria should exhibit a comparatively higher abundance of defense systems. Indeed, "habitat = free living" ranks within the top 19.

Additionally, aquatic environments and energy sources emerged as key factors. "Habitat = aquatic" and "energy source = photoautotroph" ranked prominently, indicating their influence on

defense system abundance (Fig. 2). Additionally, polyhabitat, filamentous shape, mobility, and bioremediation also appear to influence defense system abundance significantly.

Further PGLS analysis revealed negative correlations between symbiotic characteristics and defense system abundance, while free-living bacteria exhibited a positive association (Table 2). While some correlations may not achieve statistical significance, all other ecological characteristics positively correlate with the abundance of defense systems (Table 2). Notably, bacteria inhabiting multiple habitats, exhibiting motility, or being involved in environmental bioremediation tend to possess a higher abundance of defense systems. Given their exposure to complex or fluctuating environments, these bacteria are more susceptible to encountering invasive elements such as phages. Another intriguing finding is that filamentous bacteria harbor significantly more defense systems, likely due to their filamentous morphology providing increased sites for phage attachment.

Ecological Factors Impacting Bacterial Growth Rates: Insights from Endosymbiosis and Habitat Dynamics

We investigate whether ecological factors influencing defense system abundance affect bacterial growth rates. Using PGLS regression, we correlated minimal doubling time with overall maximal confidence scores for significant ecological features.

Endosymbiosis and symbiosis with insects differ in their effects on bacterial growth rates compared to symbiosis with mammals (Table 3). "Known habitats = insect endosymbiont," "Known habitats = endosymbiont," and "Host = insects general" exhibit significantly positive correlations with minimal doubling time. Conversely, features related to symbiosis with mammals show negative correlations, albeit not statistically significant. Ambiguous features like "Biotic relationship = symbiotic" show a significant positive correlation, while others like "Ecosystem = host associated," "Known habitats = host," and "Habitat = host associated" exhibit negative correlations.

Aquatic habitat and energy source features positively correlate with minimal doubling time (Table 3), suggesting slower growth rates in bacteria residing in aquatic environments and utilizing natural energy sources. Additionally, filamentous morphology correlates with slower growth rates, while bacteria inhabiting multiple habitats tend to grow faster (Table 3).

Exploring Growth-Defense Relationships Across Ecological Niches

In our analysis utilizing the ProTraits database, each species received an overall maximal confidence score indicating the likelihood of possessing specific ecological features. We categorized species into presence or absence groups based on a threshold of 0.5, compiling species counts for the top 30 influential ecological features affecting defense system abundance. Features with fewer than 20 species were filtered out.

To investigate potential growth-defense trade-offs within these ecological niches, we conducted PGLS regression analysis, correlating bacterial minimal doubling time with defense system abundance (Table 4). Highly significant correlations ($P < 0.01$) were primarily observed in two niches: "Habitat = free living" and "Mobility," with extensive sample sizes comprising hundreds of bacterial species.

Among the features related to symbiosis, significant correlations between growth and defense were identified in three features, each representing samples of over 400 species: "Ecosystem = host associated," "Known habitats = host," and "Habitat = host associated." Accounting for phylogenetic data, closely related lineages with similar characteristics were treated as nearly one effective sample. Thus, the effective sample size in phylogenetic comparative analysis was substantially lower than the total number of analyzed lineages. Consequently, the conventional statistical rule of thumb regarding small and large samples ($n = 30$) should not be applied in phylogenetic comparative studies (Hu et al. 2022).

Nevertheless, the findings presented in Table 4 suggest the presence of not only a trade-off but also a trade-up between bacterial growth and defense.

Contextual Dependency in Growth-Defense Correlations Across Bacterial Environments

In Table 4, diverse correlations emerge between bacterial growth and defense across different environments. Positive correlations are observed in some bacteria within specific habitats, while

negative correlations are apparent in others; meanwhile, no correlation is discernible in some instances. Consequently, within a large sample, the prevalence of each bacterial type dictates the observed correlation pattern: a predominance of bacteria exhibiting positive correlations yields a positive overall correlation, a balance of the three types results in no discernible correlation, and a prevalence of bacteria with negative correlations leads to an overall negative correlation. Thus, the significance of analyzing growth-defense correlations on a global scale diminishes.

Nevertheless, we conducted PGLS regression analysis to explore the growth-defense correlation in our dataset of 3532 samples, yielding statistically insignificant results ($P = 0.398$). Despite the lack of statistical significance, our analysis underscores the complex and context-dependent nature of the relationship between bacterial growth rates and defense mechanisms across diverse ecological contexts.

Discussion

The analysis of bacterial defense mechanisms, whether assessed by the presence or absence of defense systems, or their abundance, yielded notable findings: ecological factors associated with symbiosis and endosymbiosis emerged as the primary influencers of bacterial defenses (Figs. 1, 2). Furthermore, phylogenetic regression analyses demonstrated a trend wherein symbiotic or endosymbiotic bacteria tend to exhibit fewer defense systems or lack them altogether (Table 1, 2). This observation aligns with prior findings noting the scarcity of defense systems among insect endosymbionts (Makarova et al. 2013; Koonin et al. 2017). Our study unveils statistically significant associations between symbiosis/endosymbiosis and defense mechanisms, suggesting that such bacteria have reduced the necessity for robust defensive strategies. Endosymbiotic bacteria, residing within host cells, enjoy physical insulation from aggressive phages. Similarly, evidence suggests that many symbiotic bacteria are spatially separated from such phages. Notably, the gut microbiome represents a well-studied example of symbiotic bacteria, comprising the majority of human symbionts (Sender et al. 2016a; Sender et al. 2016b). Within the mammalian small intestine and colon, specialized epithelial cells secrete a mucosal layer (Donaldson et al. 2016), which serves as a

formidable barrier against the spread and access of phage populations, providing a refuge for bacteria (Guerin and Hill 2020; Lourenço et al. 2020; Shkoporov et al. 2022). Consequently, the phage-bacterial ratios in the human gut are significantly lower compared to other ecosystems, often ranging from 1:1 to 0.1:1 as opposed to 10:1(Shkoporov and Hill 2019).

However, the impact of endosymbiosis on bacterial growth rates in insects differs significantly from symbiosis in mammals (Table 3). The proliferation of endosymbionts is not solely determined by their inherent ability to replicate but is also influenced by regulatory constraints imposed by their hosts (Whittle et al. 2021; Whittle et al. 2023). In contrast, the gut presents a highly complex environment for bacteria. Nutrient levels fluctuate considerably on a daily basis, and the presence of antimicrobial molecules and oxygen secreted by the epithelium strongly restricts potential microbial inhabitants (Donaldson et al. 2016), creating a gradient of suitability for bacterial growth from the lumen to the mucosa. Bacteria residing in isolated microhabitats such as crypts may be tightly controlled by the host epithelium, while those dispersed into the gut lumen could proliferate in response to nutrient availability. We hypothesize that the overall impact of gut symbiosis on bacterial growth rate depends on the distribution of bacteria across different microhabitats.

Due to insufficient sample size, no statistical results have been yielded regarding the relationship between growth and defense in endosymbiotic bacteria. However, a weak negative correlation between minimal doubling time and the abundance of defense systems was observed in symbiotic bacteria (Table 4). This suggests that symbiotic bacteria with robust defense mechanisms tend to exhibit rapid growth. Rather than a trade-off, there appears to be a potential trade-up between growth and defense in symbiotic bacteria. We speculate that, despite potential fluctuations in nutrient levels, hosts typically provide a eutrophic environment for symbiotic bacteria, where energy and resources are not limiting factors, thus rendering the trade-off between growth and defense unnecessary. The gut lumen may favor *r*-selected bacteria, which prioritize rapid growth and reproduction in unstable environments.

Several ecological features, such as free-living habitat, phototrophic lifestyle, occupation of multiple habitats, motility, filamentous morphology, and involvement in bioremediation, have been found to be positively associated with the abundance of defense systems (Table 2). The trade-off

between growth and defense has been statistically demonstrated in free-living and motile bacteria. These findings suggest that certain ecological niches may favor the allocation of resources towards defense mechanisms, potentially at the expense of growth rates, in order to cope with environmental challenges such as predation by bacteriophages. The positive association between defense system abundance and ecological features like free-living habitat and motility implies that bacteria thriving in diverse and dynamic environments may require heightened defense strategies to survive and proliferate. However, further research is needed to explore the specific mechanisms underlying these associations and to determine how ecological factors influence the evolution of bacterial defense strategies.

Supplementary Information

The supplementary material is submitted alongwith the main manuscript online.

Acknowledgements

This study was funded by the National Natural Science Foundation of China (grant number 31671321). We are grateful to Tiao-Hao Zhao for his help in machine learning techniques.

Author Contributions

DKN conceived the research hypothesis; ZLL and JL analyzed the data. ZLL and DKN drafted the manuscript; all authors contributed equally to the final version and reviewed the manuscript.

Data Availability

The datasets analyzed during the current study are available in the Supplementary Files.

References

Bernheim A, Sorek R (2020) The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol 18:113-119

Bohannan BJM, Kerr B, Jessup CM, Hughes JB, Sandvik G (2002) Trade-offs and coexistence in microbial microcosms. Antonie van Leeuwenhoek 81:107-115

Bohannan BJM, Lenski RE (2000) Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol Lett 3:362-377

Brbić M, Piškorec M, Vidulin V, Kriško A, Šmuc T, Supek F (2016) The landscape of microbial phenotypic traits and associated genes. Nucleic Acids Res 44:10074-10090

Buckling A, Wei Y, Massey RC, Brockhurst MA, Hochberg ME (2005) Antagonistic coevolution with parasites increases the cost of host deleterious mutations. Proc R Soc B 273:45-49

Burmeister AR, Fortier A, Roush C, Lessing AJ, Bender RG, Barahman R, Grant R, Chan BK, Turner PE (2020) Pleiotropy complicates a trade-off between phage resistance and antibiotic resistance. Proc Natl Acad Sci USA 117:11207-11216

Burmeister AR, Turner PE (2020) Trading-off and trading-up in the world of bacteria-phage evolution. Curr Biol 30:R1120-R1124

Castledine M, Padfield D, Sierocinski P, Soria Pascual J, Hughes A, Mäkinen L, Friman V-P, Pirnay J-P, Merabishvili M, de Vos D, Buckling A (2022) Parallel evolution of Pseudomonas aeruginosa phage resistance and virulence loss in response to phage treatment in vivo and in vitro. eLife 11:e73679

Chen Z-L, Guo H-J, Niu D-K (2023) Dependent variable selection in phylogenetic generalized least squares regression analysis under Pagel's lambda model. bioRxiv:2023.2005.2021.541623

Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246-W251

Debastiani VJ, Bastazini VAG, Pillar VD (2021) Using phylogenetic information to impute missing functional trait values in ecological databases. Ecol Inform 63:101315

Donaldson GP, Lee SM, Mazmanian SK (2016) Gut biogeography of the bacterial microbiota. Nat Rev Microbiol 14:20-32

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1-15

Ferenci T (2016) Trade-off mechanisms shaping the diversity of bacteria. Trends Microbiol 24:209-223

Garland Jr T (2014) Trade-offs. Curr Biol 24:R60-R61

Garland T, Downs CJ, Ives AR (2022) Trade-offs (and constraints) in organismal biology. Physiol Biochem Zool 95:82-112

Garland T, Jr., Dickerman AW, Janis CM, Jones JA (1993) Phylogenetic analysis of covariance by computer simulation. Syst Biol 42:265-292

Georjon H, Bernheim A (2023) The highly diverse antiphage defence systems of bacteria. Nat Rev Microbiol 21:686-700

Guerin E, Hill C (2020) Shining Light on Human Gut Bacteriophages. Front Cell Infect Microbiol 10

Hall AR, De Vos D, Friman V-P, Pirnay J-P, Buckling A (2012) Effects of sequential and simultaneous applications of bacteriophages on populations of *Pseudomonas aeruginosa* in vitro and in wax moth larvae. Appl Environ Microbiol 78:5646-5652

Herms DA, Mattson WJ (1992) The dilemma of plants: to grow or defend. Q Rev Biol 67:283-335

Ho LST, Ane C (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst Biol 63:397-408

Hu E-Z, Lan X-R, Liu Z-L, Gao J, Niu D-K (2022) A positive correlation between GC content and growth temperature in prokaryotes. BMC Genomics 23:110

Ives AR, Garland T, Jr. (2010) Phylogenetic logistic regression for binary dependent variables. Syst Biol 59:9-26

Jessup CM, Bohannan BJM (2008) The shape of an ecological trade-off varies with environment. Ecol Lett 11:947-959

Koonin EV, Makarova KS, Wolf YI (2017) Evolutionary genomics of defense systems in archaea and bacteria. Annu Rev Microbiol 71:233-261

Koskella B, Taylor TB, Bates J, Buckling A (2011) Using experimental evolution to explore natural patterns between bacterial motility and resistance to bacteriophages. ISME J 5:1809-1817

Lennon JT, Khatana SAM, Marston MF, Martiny JBH (2007) Is there a cost of virus resistance in marine cyanobacteria? ISME J 1:300-312

Lenski RE (1988) Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. Evolution 42:425-432

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2:18-22

Liu Z-L, Hu E-Z, Niu D-K (2023) Investigating the relationship between CRISPR-Cas content and growth rate in bacteria. Microbiol Spectr 11:e03409-03422

López-García P, Gutiérrez-Preciado A, Krupovic M, Ciobanu M, Deschamps P, Jardillier L, López-Pérez M, Rodríguez-Valera F, Moreira D (2023) Metagenome-derived virus-microbe ratios across ecosystems. ISME J 17:1552-1563

Lourenço M, Chaffringeon L, Lamy-Besnier Q, Pédron T, Campagne P, Eberl C, Bérard M, Stecher B, Debarbieux L, De Sordi L (2020) The spatial heterogeneity of the gut limits predation and fosters coexistence of bacteria and bacteriophages. Cell Host Microbe 28:390-401.e395

Madin JS, Nielsen DA, Brbic M, Corkrey R, Danko D, Edwards K, Engqvist MKM, Fierer N, Geoghegan JL, Gillings M, Kyrpides NC, Litchman E, Mason CE, Moore L, Nielsen SL, Paulsen IT, Price ND, Reddy TBK, Richards MA, Rocha EPC, Schmidt TM, Shaaban H, Shukla M, Supek F, Tetu SG, Vieira-Silva S, Wattam AR, Westfall DA, Westoby M (2020) A synthesis of bacterial and archaeal phenotypic trait data. Scientific Data 7:170

Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas C, White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18:67-83

Makarova KS, Wolf YI, Koonin EV (2013) Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res 41:4360-4377

Moulton-Brown CE, Friman V-P (2018) Rapid evolution of generalized resistance mechanisms can constrain the efficacy of phage–antibiotic treatments. Evol Appl 11:1630-1641

Pančić M, Kiørboe T (2018) Phytoplankton defence mechanisms: traits and trade-offs. Biol Rev 93:1269-1303

Quance MA, Travisano M (2009) Effects of temperature on the fitness cost of resistance to bacteriophage T4 in *Escherichia coli*. Evolution 63:1406-1416

Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298-D299

Roff DA, Fairbairn DJ (2007) The evolution of trade-offs: where are we? J Evol Biol 20:433-447

Rostøl JT, Marraffini L (2019) (Ph)ighting phages: how bacteria resist their parasites. Cell Host Microbe 25:184-194

Roux S, Brum JR (2023) Counting dots or counting reads? Complementary approaches to estimate virus-to-microbe ratios. ISME J 17:1521-1522

Santos T (2018) Phylogenetic eigenvectors regression and phylogentic signal-representation curve R Cran Project

Sender R, Fuchs S, Milo R (2016a) Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell 164:337-340

Sender R, Fuchs S, Milo R (2016b) Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 14:e1002533

Shkoporov AN, Hill C (2019) Bacteriophages of the human gut: the "known unknown" of the mcrobiome. Cell Host Microbe 25:195-209

Shkoporov AN, Turkington CJ, Hill C (2022) Mutualistic interplay between bacteriophages and bacteria in the human gut. Nat Rev Microbiol 20:737-749

Stekhoven DJ, Bühlmann P (2011) MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 28:112-118

Stern A, Sorek R (2011) The phage-host arms race: Shaping the evolution of microbes. BioEssays 33:43-51

Suttle CA (2007) Marine viruses - major players in the global ecosystem. Nat Rev Microbiol 5:801-812

Symonds MRE, Blomberg SP (2014) A primer on phylogenetic generalised least squares. In: Garamszegi LZ (ed) Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 105-130

Tesson F, Bernheim A (2023) Synergy and regulation of antiphage systems: toward the existence of a bacterial immune system? Curr Opin Microbiol 71:102238

Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A (2022) Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun 13:2561

Thingstad TF (2022) Competition-defense trade-offs in the microbial world. Proc Natl Acad Sci USA 119:e2213092119

Weissman JL, Hou S, Fuhrman JA (2021) Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proc Natl Acad Sci USA 118:e2016810118

Weissman JL, Laljani RMR, Fagan WF, Johnson PLF (2019) Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. ISME J 13:2589-2602

Whittle M, Barreaux AMG, Bonsall MB, Ponton F, English S (2021) Insect-host control of obligate, intracellular symbiont density. Proc R Soc B 288:20211993

Whittle M, Bonsall MB, Barreaux AMG, Ponton F, English S (2023) A theoretical model for host-controlled regulation of symbiont density. J Evol Biol 36:1731-1744

Winter C, Bouvier T, Weinbauer MG, Thingstad TF (2010) Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited.

Microbiol Mol Biol Rev 74:42-57
Wommack KE, Colwell RR (2000) Virioplankton: Viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69-114
Yang JW, Chang FH, Yeh YC, Tsai AY, Chiang KP, Shiah FK, Gong GC, Hsieh CH (2023) Trade-offs between competitive ability and resistance to top-down control in marine microbes. Msystems 8:e01017-01022

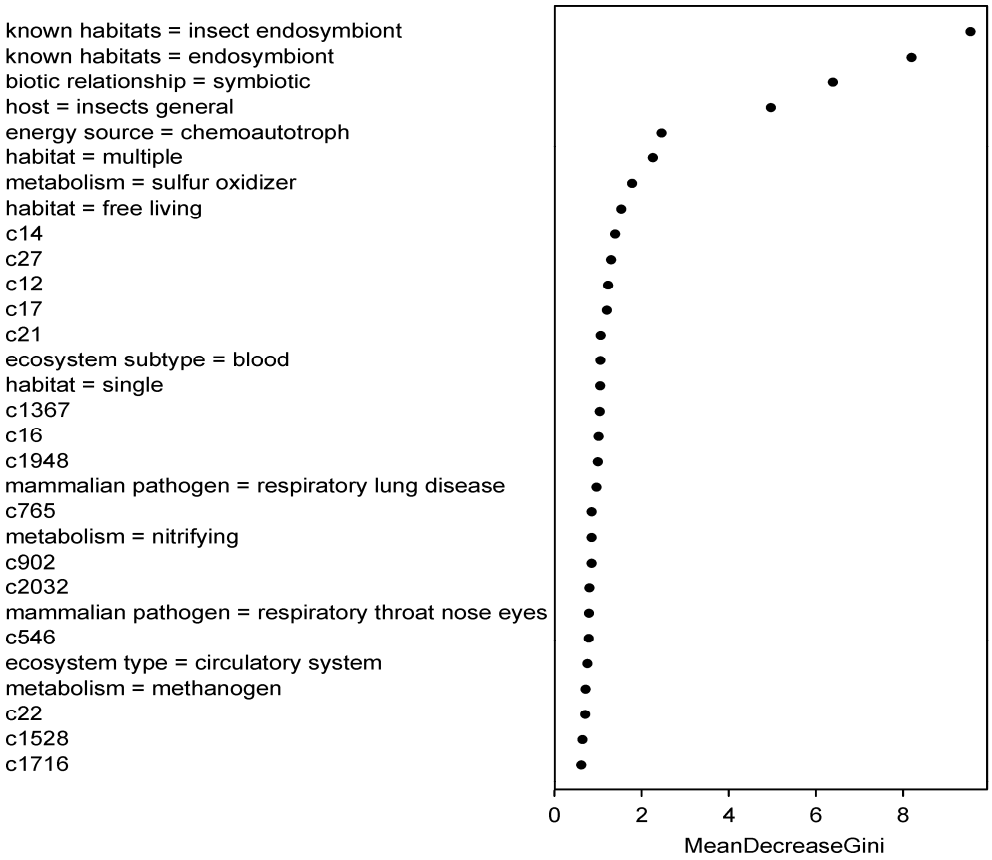

Figure 1. Importance ranking of ecological features for defense system presence/absence in bacteria. The vertical axis represents the names of ecological features, with higher ranks indicating a greater impact on the presence or absence of the defense system. MeanDecreaseGini, displayed on the horizontal axis, quantifies feature importance, with higher values indicating greater significance. The figure displays the top thirty features in the model. Features starting with the letter "C" represent eigenvectors representing phylogenetic relationships.

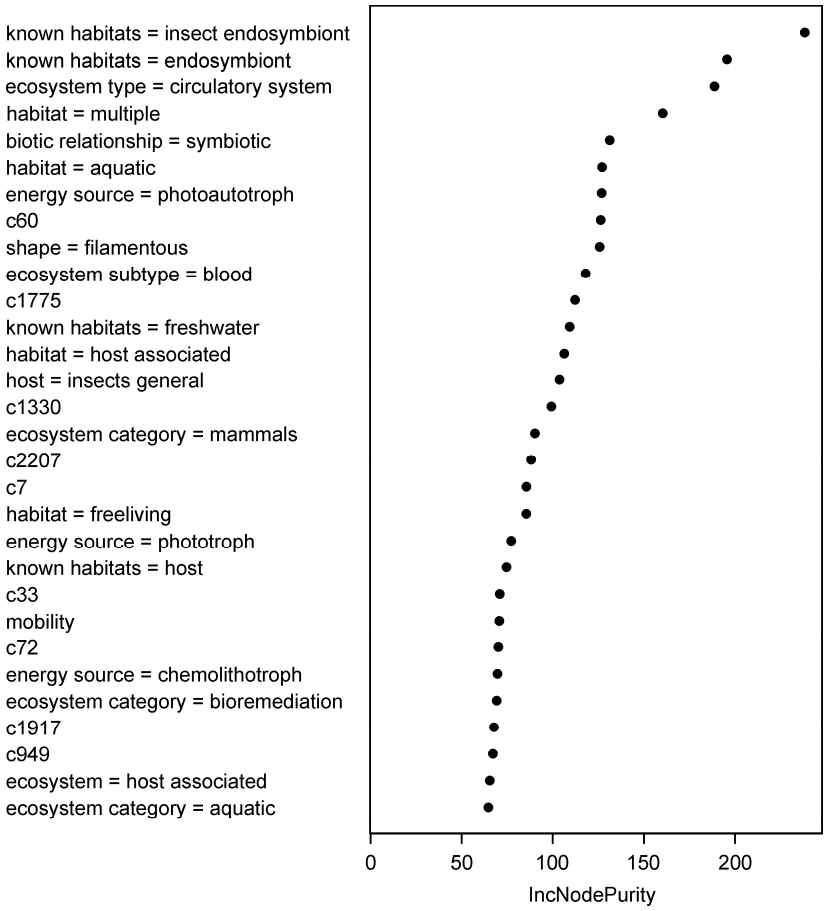

Figure 2. Importance ranking of ecological features for defense system abundance in bacteria. The vertical axis represents the names of ecological features, with higher ranks indicating a greater impact on the abundance of the defense system. IncNodePurity, displayed on the horizontal axis, quantifies feature importance, with higher values indicating greater significance. The figure displays the top thirty features in the model. Features starting with the letter "C" represent eigenvectors representing phylogenetic relationships.

Table 1. Phylogenetic logistic regression of defense system presence/absence on bacterial habitats

Dependent variable ~ independent variable	Slope	P	$P_{adjusted}$
Presence/absence of defense ~ Known habitats = insect endosymbiont	–15.785	4.634×10^{-14}	1.854×10^{-13}
Presence/absence of defense ~ Known habitats = endosymbiont	0.286	0.452	0.452
Presence/absence of defense ~ Biotic relationship = symbiotic	–7.344	1.54×10^{-10}	3.080×10^{-10}
Presence/absence of defense ~ Host = insects general	–8.215	3.0×10^{-9}	4.0×10^{-9}

The default parameter logistic MPLE was used in the phylogenetic logistic regression of the relationships. In the regression analysis, the presence of defense systems was assigned as 1 and the absence was assigned as 0. The habitats are measured by the overall maximal confidence score described in Materials and Methods section. The ratio of samples with defense systems to those without related systems was 962 vs 82. The $P_{adjusted}$ value represents the corrected P-value computed using the Benjamini & Hochberg method.

Table 2. PGLS regressions of ecological features on the defense system abundance in bacteria

	Slope	P	$P_{adjusted}$
Known habitats = insect endosymbiont	–0.001	0.001	**0.004**
Known habitats = endosymbiont	–0.001	4.4×10^{-4}	**0.002**
Biotic relationship = symbiotic	–0.002	0.022	**0.031**
Ecosystem type = circulatory system	–0.001	0.019	**0.029**
Ecosystem subtype = blood	–0.001	0.068	0.084
Habitat = host associated	–0.002	0.002	**0.005**
Host = insects general	–0.002	0.002	**0.005**
Ecosystem category = mammals	–0.002	0.015	**0.024**
Known habitats = host	–0.002	0.008	**0.017**
Ecosystem = host associated	–0.002	0.011	**0.019**
Habitat = free living	0.002	0.007	**0.016**
Habitat = aquatic	0.001	0.054	0.071
Known habitats = freshwater	0.001	0.190	0.210
Ecosystem category = aquatic	0.001	0.284	0.284
Energy source = photoautotroph	0.001	0.251	0.264
Energy source = phototroph	0.003	8.5×10^{-6}	$\mathbf{9.0\times10^{-5}}$
Energy source = chemolithotroph	0.001	0.105	0.123
Habitat = multiple	0.002	0.001	**0.004**
Shape = filamentous	0.002	3.8×10^{-6}	$\mathbf{8.0\times10^{-5}}$

Mobility	0.002	0.010	**0.019**
Ecosystem category = bioremediation	0.002	3.9×10^{-5}	$\mathbf{2.7\times10^{-4}}$

Due to the weaker phylogenetic signal of defense system abundance in bacteria compared to ecological features (Table S1), it was utilized as the independent variable in the PGLS regression analysis. Ecological features were measured using the overall maximal confidence score described in the Materials and Methods section. A total of 1044 bacterial species were included in these regressions. The P_{adjusted} value represents the corrected P-value computed using the Benjamini & Hochberg method. For readability, we have highlighted P_{adjusted} values less than 0.05 in bold font. PGLS, phylogenetic generalized least-squares. AIC, Akaike information criterion.

Table 3. PGLS regressions of bacterial minimal doubling time on ecological features

	Slope	P	$P_{adjusted}$
Known habitats = insect endosymbiont	1.809	1.0×10^{-5}	$\mathbf{3.0\times10^{-5}}$
Known habitats = endosymbiont	0.799	0.004	**0.006**
Host = insects general	1.07	3.9×10^{-9}	$\mathbf{4.1\times10^{-8}}$
Biotic relationship = symbiotic	0.609	3.3×10^{-5}	$\mathbf{8.7\times10^{-5}}$
Ecosystem = host associated	–0.859	8.4×10^{-8}	$\mathbf{5.9\times10^{-7}}$
Known habitats = host	–0.625	9.2×10^{-5}	$\mathbf{1.9\times10^{-4}}$
Habitat = host associated	–0.478	0.004	**0.006**
Ecosystem type = circulatory system	–0.197	0.325	0.341
Ecosystem subtype = blood	–0.27	0.097	0.113
Ecosystem category = mammals	–0.111	0.503	0.503
Habitat = free living	0.253	0.089	0.110
Habitat = aquatic	0.907	9.7×10^{-7}	$\mathbf{3.4\times10^{-6}}$
Known habitats = freshwater	0.693	6.7×10^{-5}	$\mathbf{1.6\times10^{-4}}$
Ecosystem category = aquatic	0.92	7.7×10^{-7}	$\mathbf{3.2\times10^{-6}}$
Energy source = photoautotroph	1.235	7.5×10^{-7}	$\mathbf{3.2\times10^{-6}}$
Energy source = phototroph	0.422	0.035	**0.049**
Energy source = chemolithotroph	1.518	2.5×10^{-12}	$\mathbf{5.3\times10^{-11}}$
Habitat = multiple	–0.656	2.2×10^{-4}	$\mathbf{4.2\times10^{-4}}$
Shape = filamentous	0.845	0.002	**0.004**

Mobility	–0.305	0.081	0.106
Ecosystem category = bioremediation	0.293	0.311	0.341

Since the minimal doubling time of bacteria exhibits a stronger phylogenetic signal than the ecological features (Table S1), we employed it as the dependent variable in the PGLS regression analysis. Ecological features were measured using the overall maximal confidence score described in the Materials and Methods section. A total of 1044 bacterial species were included in these regressions. The $P_{adjusted}$ value represents the corrected P-value computed using the Benjamini & Hochberg method. For readability, we have highlighted $P_{adjusted}$ values less than 0.05 in bold font. PGLS, phylogenetic generalized least-squares. AIC, Akaike information criterion.

Table 4. PGLS regressions of bacterial minimal doubling time on the abundance of defense systems

	n	Slope	*P*
Host = insects general	30	0.092	0.478
Biotic relationship = symbiotic	98	–0.009	0.210
Ecosystem = host associated	499	–0.015	**0.028**
Known habitats = host	518	–0.016	**0.021**
Habitat = host associated	434	–0.017	**0.020**
Ecosystem type = circulatory system	52	–0.012	0.194
Ecosystem subtype = blood	92	–0.018	0.104
Ecosystem category = mammals	81	–0.019	0.129
Habitat = free living	605	0.013	**0.008**
Habitat = aquatic	194	0.007	0.265
Known habitats = freshwater	147	0.001	0.901
Ecosystem category = aquatic	302	0.007	0.243
Energy source = photoautotroph	20	0.028	**0.023**
Energy source = chemolithotroph	22	0.01	0.524
Mobility	410	0.02	**0.002**
Habitat = multiple	206	0.01	0.264
Shape = filamentous	22	–0.001	0.837

Since the minimal doubling time of bacteria exhibits a stronger phylogenetic signal than the abundance of defense systems (Table S1), we used it as the dependent variable in the PGLS regression analysis. We discarded the ecological features where less than 20 bacterial samples were

obtained. We did adjust the *P*-values in this table using the Benjamini & Hochberg method because the multiple regressions were not performed on the same dataset. For readability, we have highlighted *P* values less than 0.05 in bold font. PGLS, phylogenetic generalized least-squares. AIC, Akaike information criterion.